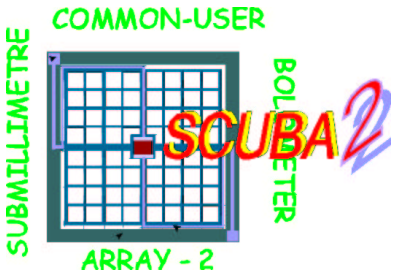

SCUBA-2 DR Pipeline Project Office
University of British Columbia
6224 Agricultural Road
Vancouver, British Columbia
CANADA
V6T 1Z1

Tel: +1-604-822-2211
Fax: +1-604-822-5324
Email: jmolnar@ubc.ca
WWW: http://scuba2.jach.hawaii.edu

Document Title: SCUBA-2 Data Simulation and Pipeline Testing

Document Number: SC2/SOF/S210/003

Issue: 1.19

Date: 2005/05/28

Document Prepared By:	Douglas Scott	Signature and Date:	2003-11-29
Document Approved By:	Janos Molnar	Signature and Date:	2003-11-30
Document Released By:	Janos Molnar	Signature and Date:	2003-11-30

Contents

1 Introduction

The SCUBA-2 Data Reduction Software will need to be fully tested with simulated data. This must include the full On-line, Off-line and Data Display systems.

As each module of code is written it will need to be tested for functionality and speed. It will then need to be tested as part of the Pipeline. And finally there will need to be tests of the entire Pipeline, running each Recipe, with a variety of sorts of input data. The simplest will be random numbers, with the same format as real data. But it will be necessary to test the data fully on simulated data which are as realistic as possible.

For all of the Software, tests with random numbers (or other ideal data) will provide simple checks that the Software can handle the expected data-rate and that the basic infra-structure works. But ultimately the DR Software will need to be tested on simulated data which is as similar to the real data as possible, which implies a requirement on the SCUBA-2 data simulator.

The On-line system needs to tested in the following ways:

1. Test that it fulfills all of the functionality requirements specified in the Data Reduction Software Requirements document.[15]
2. Test that it is able to keep up with the full data rate for all of the observing modes and Recipes. ***Req. PR5,MR3***
3. Tests of robustness under different possible observing conditions, changes between modes, etc., to check that it requires minimal human intervention. ***Req. GR7***

The Off-line system needs to be tested in the following ways:

1. Test that it has the full functionality as required in the Data Reduction Software Requirements document.[15]
2. Test that it works at a number of different institutions. ***Req. OR8***
3. Tests that it can fully duplicate the DREAM and STARE reduction steps which are normally carried out by the Data Acquisition System. ***Req. MR2***
4. Check that mosaicing of large images works, which can be done with simulated data. ***Req. OR5***

The Data Display system[3] needs to be tested in the following ways:

1. Test the functionality and speed of the Stripchart facility. ***Req. DR12–DR19***
2. Test that on-line 'movie' mode can display at a reasonable rate, and investigate the usefulness of this mode. ***Req. DR5***
3. Test that Display output images are displayed at a sufficient rate and are of sufficient quality to be useful. ***Req. DR7–DR9***

There are also some general tests that need to performed throughout the development of the Software:

1. Check that the Software runs on Solaris machines, and (currently at a lower priority) on MACs. The Software is being developed on computers running Linux, but it will also be necessary to port to other versions of Linux. ***Req. GR1,GR20***
2. Test that the Software runs reliably and can recover from different types of failure. ***Req. GR7,GR8***
3. Check that the 'data expander' (intended to write out data in a usable format), is useful to the user community, and that users can successfully include their own algorithms when they re-run the Software at their own institutions. ***Req. GR15,OR7***

2 Simulations

Testing of the Software requires the generation of simulated data. These simulations need to be as realistic as possible. They will build on what has already been developed for the Data Acquisition effort at ATC.[6,9] The development of the simulator is the formal responsibility of the DA-Software work package at the ATC. However, running the simulator in order to test the DR Software is clearly part of the DR Software work. In addition, investigation of SCAN mode[17,16] (which is not reduced in the DA computers) is naturally part of the DR work package. The simulation of realistic SCUBA-2 data is an ongoing process, which will add complications in a staged way, with the final result being the generation of test data which are as similar as possible to real data. So, although the responsibility for developing the simulator rests with ATC, it is expected that work on SCAN mode will help define improved features of the simulations. ***Req. TR5***

The simulated data need to be in the correct format, as specified in the Interface Control Document.[2] Before the Software is delivered, the data need to be simulated at the full rate for all of the sub-arrays, with noise sources included, and for a variety of test images. Tests of the Software can start with the simplest simulations, but eventually the simulated data need to include: ***Req. ER3–ER5***

1. Fake images containing bright point sources, faint point sources, extended emission and confusion noise.
2. Noise (white plus $1/f$) from the detectors, common-mode as well as realizations for each bolometer and correlations among rows of bolometers, and including both gain and offset variations.
3. Accurate simulated sky noise, in particular with a realistic power spectrum for total-power detectors and possible spatial variation across the array.
4. Variable sky opacity, with approximately correct statistics, and for different weather bands.
5. Cosmic ray effects.
6. Individual bad bolometers as well as entire rows.
7. Secondary mirror jitter and acceleration effects from scanning with the primary.

Several of these effects have already been included in the ATC Simulator,[9] although some others have not yet been considered. Some of the details remain to be investigated, for example whether the opacity variations can also come from the code which simulates the sky emission. As of the time of the DR Software CDR, little is known about the behaviour of the

real bolometer arrays. The simulations will certainly need to be revised as more information is gained about likely instrument performance. New effects (for example unwanted microphonics, synchronous sources of noise or other unexpected correlations) will also need to be included at levels which may be estimated for the instrument as it is built and tested.

3 Reduction algorithms

Algorithm development is a crucial step which needs to be completed before SCUBA-2 is commissioned. There needs to be investigation of how well images are obtainable using the STARE, DREAM and SCAN modes. Which steps are needed in each of the procedures will depend in practice on the noise properties of the data. Some background discussion can be found in SC2/SOF/S200/015,[5] SC2/ANA/S210/001[14] and SC2/ANA/S210/006.[16] Documentation will be worked on concurrently with coding, and full documentation will be delivered along with the Software. ***Req. TR1***

For the STARE mode, there needs to be some effort to understand what simple reduction steps might be taken to improve the image under realistic observing conditions. This is the simplest mode, and hence there are the least number of possibilities for what can be carried out on the individual frames. However, there still needs to be an investigation of removing planes rather than the array average, as well as the possibility of considering higher-order functions or more local averages. The reduction procedure also needs to handle the jiggle-pattern necessitated to fully-sample the $450\,\mu$m array, as well as the possibility of bad bolometers, which will also need to be tested.

The DREAM mode[12] is the procedure for which the most investigative work has already been done.[10,18,7,8] However, there still needs to be investigation of the optimization of the pattern, the speed of the secondary motion and details of the reduction steps. These will all have to be done in association with increasingly realistic simulations.

SCAN is the mode which still requires the most investigative effort, and this is the responsibility of the DR Software work package. Although some work has already been carried out,[4] this is of a very preliminary nature. However, the Project is not working in a vacuum here, and there are useful lessons to be learned from other modern bolometric array detectors, particularly those which operate in total power mode (ACBAR, APEX, BLAST, BOLOCAM and SHARC-II for example). The basic plan is to follow the approach of the CRUSH package[11] for SHARC-II, in which models are fit one at a time and iterated if necessary. Details of the specific Primitives, the iterative procedure and the number of different available Recipes, all still remain to be defined. ***Req. TR3***

The best scan pattern and scanning speed still need to be determined,[17] in addition to the map-making approach. This task is the responsibility of the DR Software work package. There are simple techniques, where baselines along each scan and other systematics are removed one by one, then more complex methods involving matrix inversion to solve for everything at once, in order to distinguish noise in the timestream from structure on the sky. It is possible that the best strategy will be weather-dependent, or depend on the size of the region or the nature of the astronomical source being mapped. Hence it will be necessary to investigate more than one approach, so that a few separate Recipes are available for SCUBA-2 observers. ***Req. TR2*** ***Req. TR4***

4 Pipeline testing

In order to test the Pipeline there must be methods in place which simulate the DA system. These methods must result in ICD-compliant data 'appearing' on disk for the data-detection routines within ORAC-DR to identify and process. It is not necessary to generate the data at the same rate as the DA system; the data can be pre-generated and stored on disk in the correct directories. Rather, it is sufficient to write ORAC-DR flag files (also known as `.ok` files[2]) at a rate equal to that expected for each of the supported observing modes. A further simplification is also possible: the flag files themselves can be generated prior to testing and copied to `ORAC_DATA_IN` at the correct rate.

The methods differ for the Display (Quick Look) and On-line (Science) DR pipelines and each is described further below.

4.1 Display System

The Display simulator program (`qlsim`) extracts the data from the raw data files and stores them in a DRAMA parameter which ORAC-DR reads (when run with the `-loop task` option) and registers the data with the current frame object. The program reads flag files stored in `ORAC_DATA_IN` (also known as `.ok` files[2]) to determine where to locate the data and which files to read the data from.

Once the data are stored in the correct ORAC-DR frame, they are accessible to the Pipeline which can then proceed with processing them.

4.2 On-line DR

The On-line Pipeline relies on the presence of flag files to signify the presence of new data. The Pipeline reads the flag files and if new data have arrived (indicated by the addition of a new file name in *each* of the four flag files) begins to process them.

The On-line DR simulator (`drsim`) copies pre-existing flag files into the location defined by `ORAC_DATA_IN` at a rate equal to that expected for the DA system to close a given data file. (This rate is not yet defined for the real DA system although it is expected that files will be closed after an interval not exceeding the time to generate a file 1 GB in size, about 4 minutes at the SCUBA-2 data rate.) For DREAM/STARE data, it is expected that a file will be closed on disk on a shorter timescale, approximately every minute. For SCAN data, the interval may be the length of time to complete a single SCAN pattern, but this will depend on the size of the area being mapped.

For the purposes of the testing program, the interval is user-defined with a default value of 1 minute. The only other input to the program is the observation date (in `YYYYMMDD` format). The program does not know itself which files to copy and relies on information stored in a data file generate by a companion program, `createok`.

This program (`createok`) is used to create the flag files based on the observing parameters, and must be run prior to running `drsim`. The user supplies the observation date, the observation number and the number of files to be written as part of a given observation. Given

that information, the program creates the necessary flag files, as well as generating a list of the files which `drsim` will use to create the flag files in `ORAC_DATA_IN`.

Specific tests which must be done include:

- Throughput and speed tests to ensure that the Pipeline can keep up with the data rate. Staged tests can be performed, starting with simple file and command handling and ending with complete testing of an entire night's data. ***Req. PR5***
- Test that the noise is being estimated correctly for the different Recipes. Tests can include Gaussianity of blank signal-to-noise maps and comparison with the calculated noise for bright sources propagated through the reduction steps. ***Req. GR17,PR9,PR10***
- The flagging of 'bad' data needs to be tested at several stages along the road towards release of the final Software products. These tests need to include simulations of cosmic ray and other glitches, perhaps supplemented with some real (scaled) SCUBA data, or data from another total power bolometer array. Flagging pathological parts of the data is particularly important for SCAN mode.[16] ***Req. PR12***
- Test that the combination of the 4 sub-arrays is performed correctly. This requires simulating timestreams from images at least as large as the full array. ***Req. PR15***
- Use simulated data with images containing point sources to test that each of the Pipeline Recipes do not degrade the resolution by more than 10% of the beam FWHM. This will involve the use of simulations of point sources, as well as comparison among Recipes once real data are available. ***Req. MR4***
- In a similar way to the above, there will be tests of the overall astrometry errors introduced by each reduction Recipe, to ensure that they are no worse than the $450\,\mu$m resolution. This should be verifiable from tests performed on simulations. ***Req. MR5***

4.3 On-line DR

The Off-line system will use essentially identical Software to the On-line system, although some Recipes will be optimised for accuracy rather than speed, and there will be a great deal more choice. Most of the testing of functionality (apart from those already mentioned in the Introduction) and detailed Science Requirements will already be covered by tests of the On-Line system.

5 Schedule and Progress

There is an inter-dependency between the ATC, UBC and the JAC in this effort. Specifications of the sort described here are required in order for the ATC to produce simulated data in the correct format and with a sufficient level of realism. The ATC simulator is needed at various stages for testing the DR system. Work on optimising the STARE, DREAM and SCAN initial processing steps involve both the DA and DR Software. The Pipeline and Display system are needed by the ATC in order to support laboratory testing and testing of the DA system. The results from this testing are needed to improve the instrumental models in order to generate more realistic testing of the Pipeline algorithms. Development of the DR Display

system has already been driving improvements in the simulator. And lessons learned in developing algorithms will feed-back information about enhancements which might be required of the simulations.

Because of all this inter-dependency, the necessary development of increasingly realistic simulations is a complex process. There will be input from across the entire SCUBA-2 Project, and also from the general user community. Activities certainly need to be carefully coordinated in order for the simulations to be mature enough to provide a useful test of how effective the data analysis procedures are working.

Details of testing and dependencies are given in the Project Management Plan,[13] but we summarize some of the relevant key points here. Much of the basic testing activity will happen during development. Various milestones, and the tracking of the dependencies, should ensure that a working set of Data Reduction tools are delivered to the JAC before commissioning of the instrument. ***Req. TR1***

Good progress has been made on the development of the simulator (primarily for DREAM and Stare modes) with frequent communication between UBC, JAC and ATC. This has made possible throughput testing of the Pipeline software using realistic data in order to verify compliance with the 'near-real-time' requirement. ***Req. PR5***

The Stripchart[1] utility is already available for use at the JCMT, and can also be used for testing the output from the DA electronics. ***Req. DR12***

We have developed a preliminary plan for SCAN mode[17,16] which includes a summary of the examination of both simple methods (removing array averages and baselines from scans) as well as more sophisticated approaches (iterative maximum likelihood solution for the map, with the minimum amount of assumed information about the noise properties).

These reports[17,16] form the basis for beginning work over the following year, which will be continued by a new member of the DR Software team, who joins the project on 1 July 2005 and will lead the algorithm and code development for the On-line and Off-line SCAN reduction Recipes and Primitives. During this work, there will be close communication with the ATC DA-Software team (via email and regular video-cons), in order to make sure that necessary improvements in the simulator are included in future versions. ***Req. TR4*** ***Req. TR5***

Integration testing and verification of the Pipeline Software is scheduled for the first quarter of 2006, with on-site verification and commissioning support in the second quarter of 2006.

References

[1] Andy Gibb. SCUBA-2 StripChart Utility. SCUBA-2 Project SC2/SOF/S210/006, 2005.

[2] Tim Jenness. SCUBA-2 DA/DR Interface Control Document. SCUBA-2 Project SC2/SOF/IC210/01, 2003.

[3] Tim Jenness. SCUBA-2 Data Display System. SCUBA-2 Project SC2/SOF/S210/002, 2003.

[4] B. D. Kelly. SCUBA-2 scanmap simulation. SCUBA-2 Project SC2/ANA/S100/038, 2002.

[5] B. D. Kelly. Image Processing and Least-Squares Reconstructions. SCUBA-2 Project SC2/SOF/S200/015, 2003.

[6] B. D. Kelly. SCUBA-2 data simulation. SCUBA-2 Project SC2/SOF/S200/016, 2003.

[7] B. D. Kelly. SCUBA-2 noise-free DREAM simulations. SCUBA-2 Project SC2/SOF/S200/012, 2003.

[8] B. D. Kelly. DREAM Noise Simulations. SCUBA-2 Project SC2/SOF/S200/018, 2004.

[9] B. D. Kelly. SCUBA-2 Data Simulation Progress. SCUBA-2 Project SC2/SOF/S200/024, 2004.

[10] B. D. Kelly and W. D. Duncan. Application of DREAM Observing to SCUBA-2. SCUBA-2 Project SC2/ANA/S100/006, 2001.

[11] Attila Kovács. Crush: Comprehensive reduction utility for sharc-ii. URL: http://www.submm.caltech.edu/∼sharc/crush/, 2004.

[12] R. S. Le Poole and H. W. van Someren Greve. DREAM: Dutch real-time acquisition mode for SCUBA. In *Proc. SPIE, Vol. 3357, Advanced Technology MMW, Radio, and Terahertz Telescopes, Thomas G. Phillips; Ed.*, pages 638–643, July 1998.

[13] Janos Molnar. SCUBA-2 Data Reduction SW Project Management Plan. SCUBA-2 Project SC2/SOF/PM210/01, 2003.

[14] Douglas Scott. Map making in different noise regimes. SCUBA-2 Project SC2/ANA/S210/001, 2003.

[15] Douglas Scott. SCUBA-2 Data Reduction Software Requirements. SCUBA-2 Project SC2/SRE/S210/001, 2004.

[16] Douglas Scott. SCAN mode Data Reduction Strategies for SCUBA-2. SCUBA-2 Project SC2/ANA/S210/006, 2005.

[17] Alex van Engelen and Douglas Scott. SCAN Mode Strategies for SCUBA-2. SCUBA-2 Project SC2/ANA/S210/005, 2005.

[18] H. W. van Someren Greve. DREAM Algorithms for SCUBA-2. SCUBA-2 Project SC2/ANA/S100/046, 2002.